

# 深層学習による似顔絵生成ネットワークに関する研究

竹岡 大志      何 立風

## 1. はじめに

近年, 様々な SNS やアプリケーションが普及しており, プロフィール写真を登録する機会が増加したことで似顔絵を利用することがよく見られる. そうした中, 深層学習によるイラストレータの特徴を学習した似顔絵の生成に関する研究が行われている. 本研究では関連研究 (図 1) に比べて, 顔の各パーツにおいて, より独特な表現をもつ似顔絵画像の学習・自動生成 (図 2) を目指す.

そこで本研究では, 画像生成モデルである Cycle GAN のネットワークを改良し, イラストレータの描いた似顔絵と高い類似度を持つ似顔絵の生成を目的とする. その目的に対して, 画像のチャンネル情報と空間情報が与える影響を考察する.



図 1: 関連研究に用いられた似顔絵[2]



図 2: 本研究で用いた似顔絵

## 2. 関連研究

### 2.1 Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks (Cycle GAN)

J. Zhu ら[1]は画像生成モデルの一種である Cycle GAN を提案した. ペアとなる訓練データを用意する必要がないのが特徴であり, ドメインごとに画像を集めて訓練データとする. この学習ではドメインごとの画像の集合を  $X, Y$  とし,  $X$  から  $Y$ ,  $Y$  から  $X$  へそれぞれ変換する Generator を二つ ( $G: X \rightarrow Y$ ), ( $F: Y \rightarrow X$ ) を用意する. また Generator により生成された Fake 画像と訓練データの真偽を判定する Discriminator も二つ用意する. この手法では  $X$  から  $Y$  へ変換した画像  $G(X)$  を  $Y$  から  $X$  へ変換する  $F$  へ入力した画像  $F(G(X))$  と元の入力画像  $X$  の損失を表す Cycle Consistency

Loss を提案した. この Loss に加えて, Generator と Discriminator を競い合わせる Adversarial Loss と  $X$  から  $Y$  へ変換する  $G$  に  $Y$  を入力したときの損失を表す Identity Loss を用いて学習を進める.

### 2.2 敵対的生成ネットワーク (GAN) を用いたイラストレータの個性を反映した似顔絵生成手法

中島[2]は GAN を用いてイラストレータの個性を学習させて任意の顔画像に対する似顔絵を学習させたイラストレータごとのタッチで描かせるような手法を検討した. 具体的には従来手法の教師なし学習を行う Cycle GAN をもとに, 生成した似顔絵とイラストレータが描いた似顔絵のピクセル誤差である L1 Loss を学習させることで教師学習としたモデル Cyclepix を提案した. さらに Cyclepix を  $1:n$  変換に対応するべくマルチドメイン化し, Generator に顔画像とイラストレータを指すラベルを入力することで, 顔画像から特定のイラストレータの似顔絵に変換する Conditional Cyclepix として拡張させたモデルの提案も行った. 従来手法である Cycle GAN に比べ, どちらも高い類似度評価が示されたが, 更なる類似度向上が見込める結果となった.

## 3. 提案手法

### 3.1 Squeeze and Excitation Networks

Squeeze and Excitation Network (SENet) [3]は Jie Hu らによって提案された, あるネットワークと SE block を組み合わせた手法であり, SE-ResNet は ResNet と SE block を組み合わせたネットワークである. CNN は空間情報とチャンネル情報の二つを合わせて特徴を捉えている. 通常の畳み込み層では各チャンネルが均等に出力されるが, SENet ではチャンネル情報に焦点をあて, CNN の特徴マップのチャンネルを重み付けすることで情報価値の高いものを強調して比較的価値の低いものを抑え, 表現の質を上げる.

Cycle GAN では図 3 に示したように Generator のネットワークにおいて ResNet を構成する Residual block を 9 個組み込んでいる. 本研究の提案手法の一つとして各 Residual block に SE block を追加した SE-Residual block で Generator を構成する. この提案手法を CycleSE とする. Residual block を図 4, SE-Residual block を図 5 として示す.

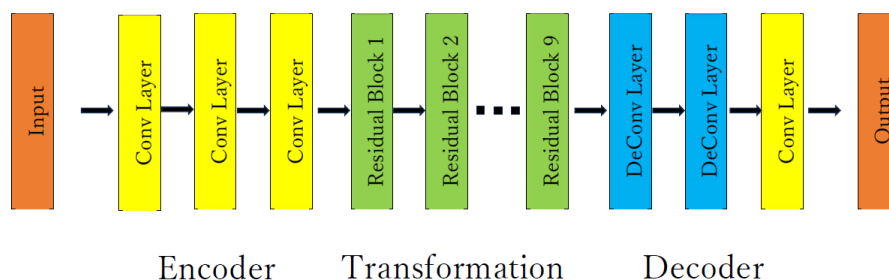


図 3: Architecture of generator

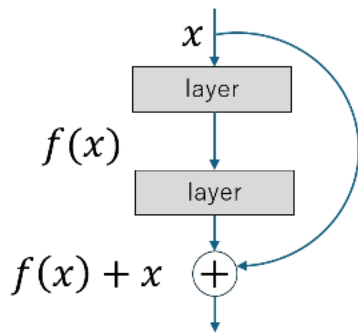


図 4 : Residual Block

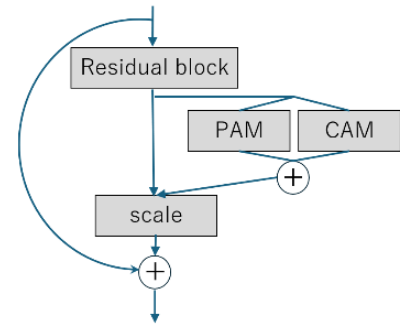


図 7 : DA-Residual Block

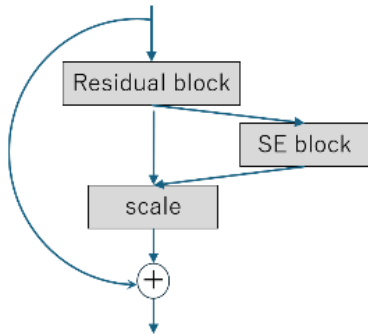


図 5 : SE-Residual Block

### 3.2 Dual Attention Network

Dual Attention Network (DANet) [4]は Jun Fu らによって提案された、セグメンテーションタスクにおける ResNet の中間層の特徴マップにおいて、画像の空間方向とチャンネル方向にそれぞれアテンションを導入した手法である。

DANet の概要図を図 6 [4]として示す。本研究の提案手法の一つとして、この DANet を構成している空間方向への Position Attention Module (図 6 : PAM), チャンネル方向への Channel Attention Module (図 6 : CAM) を各 Residual Block に追加した DA-Residual block で Generator を構成する。この提案手法を CycleDA とする。DA-Residual block を図 7 として示す。

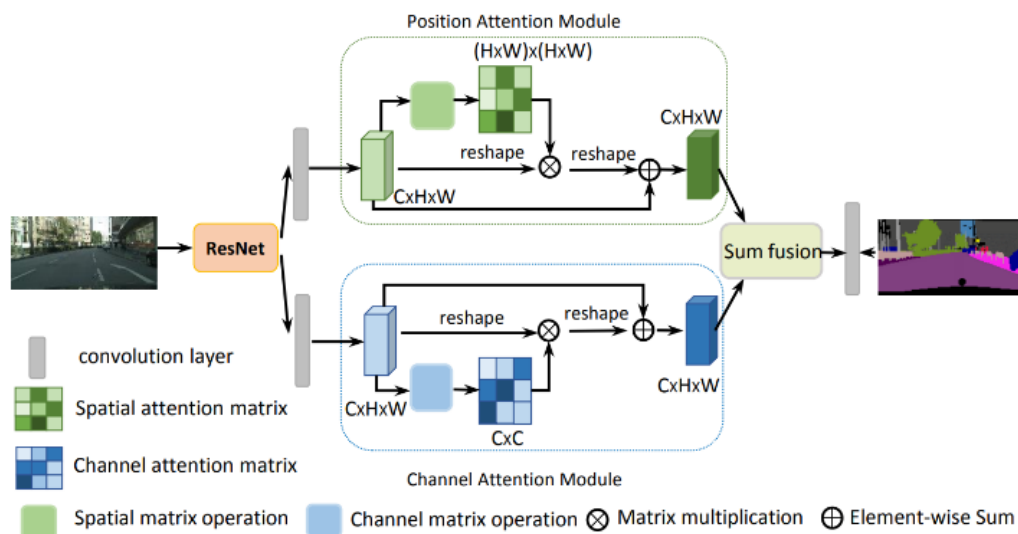


図 6 : Dual Attention

## 4 実験・評価

### 4.1 実験内容

顔画像とあるイラストレータが描いた似顔絵のペアでデータセットを作成し、訓練データが 50 枚、テストデータが 20 枚とした。

表 1 : データセット作成に用いた画像について

項目	顔画像	イラスト画像
出典元	wikipedia 首相ホームページ	読売新聞「閣僚の横顔」

損失関数に関するパラメータ  $\lambda$  の値は、Cycle GAN での初期値は  $\lambda = 10$  であったが、今回のモデルに適した  $\lambda$  値を  $10 \sim 0.1$  の中で試した結果、Cycle GAN が  $\lambda = 1.4$ , CycleSE が  $\lambda = 1.35$ , CycleDA が  $\lambda = 1.4$  とした。

本実験では Cycle GAN と、提案手法である CycleSE, CycleDA の三つを比較することで画像のチャンネル情報と空間情報が与える影響を考察する。

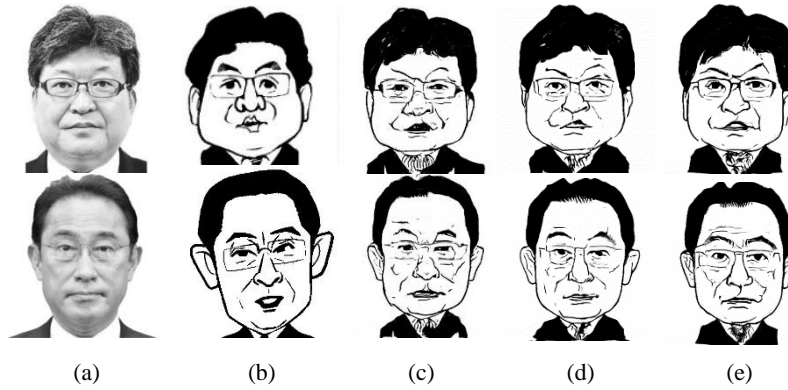


図8：訓練データにおける生成結果

#### 4.2 生成結果

図8は訓練データにおける (a) が入力画像, (b) がイラストレータ画像, (c) が Cycle GAN の生成画像, (d) が Cycle SE の生成画像, (e) が Cycle DA の生成画像である。

図9はテストデータにおける (a) が入力画像, (b) が Cycle GAN の生成画像, (c) が Cycle SE の生成画像, (d) が Cycle DA の生成画像である。

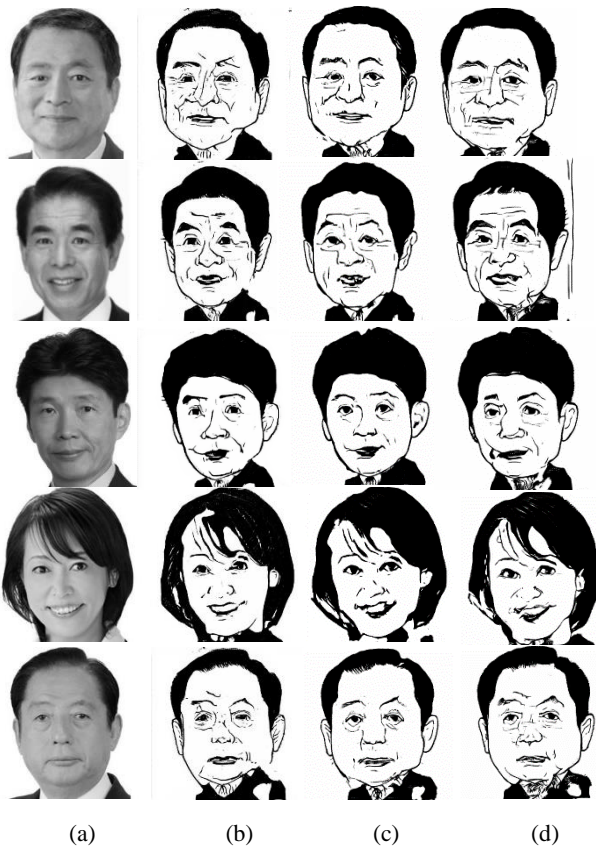


図9：テストデータにおける生成結果

#### 4.3 評価実験・考察

似顔絵の評価に関しては人の主観的な部分が多い。そのため、Cycle GAN, Cycle SE, Cycle DA の訓練データ, テストデータを被験者の評価に用いた。訓練データではイラストレータが描いた似顔絵と生成した似顔絵の類似度を評価として、被験者による評価実験を行った。イラストレータが

描いた似顔絵と生成した似顔絵において、似ていない「-2」、やや似ていない「-1」、どちらでもない「0」、やや似ている「1」、似ている「2」とした5点法を用いた。テストデータでは Cycle GAN, Cycle SE, Cycle DA の生成結果だけを見せ、訓練データに用いていないデータでの似顔絵としての完成度を評価とし、イラストレータが描いた似顔絵の雰囲気に近いかを上記と同じく5点法を用いて評価した。また、評価実験を行うため被験者を10人ほど募った。

表2：評価結果 (平均)

項目	Cycle GAN	Cycle SE	Cycle DA
訓練データ	0.401	1.099	1.169
テストデータ	0.207	1.367	0.913

表1より本研究における提案手法である Cycle SE, Cycle DA の両方で、従来手法である Cycle GAN よりも良い評価結果を得られた。

Cycle SE はチャンネル情報にアテンションするため、目における黒目と白目のような細かい領域も識別してイラストレータ画像のように表現されている。

また、Cycle DA はチャンネル情報に加えて空間情報にもアテンションするため、上記に示した Cycle SE の良さに加えて、線のタッチ感もイラストレータ画像に近いものとなっている。

しかし、どちらの提案手法でも画像によっては顔のパーツが崩れてしまうものがあつた。また、Cycle SE, Cycle DA ともに顔の細かいシワも描いてしまい、特に Cycle DA では二重顎の線や顎部分の影の影響で輪郭が崩れてしまう画像が見られた。

これらを踏まえて、評価実験における Cycle SE と Cycle DA の結果を見てみると、訓練データの結果からイラストレータ画像に対して類似度が高いのは Cycle DA である。これは目のような細かい領域に加えて線のタッチ感の表現が大きく影響していることが考えられる。しかし、テストデータの結果では Cycle SE の方が良い結果となっている。これは似顔絵としての完成度は Cycle SE の方が高いことを示しており、顔のシワや輪郭の精度が評価結果に影響していると考えられる。

このことから、チャンネル情報と空間情報にアテンションすることで似顔絵としての類似度は高くなるが、空間情報のアテンションにより、余分な情報まで反映してしまうため、完成度は低くなるのが分かった。そのため、空間情報のアテンションで余分な情報は反映しないような制約が必要だと考えている。

## 5 おわりに

今後は4章の考察にも挙げたように学習により反映された顔の細かいシワなどを除去していく必要がある。また、画像によっては顔の輪郭やパーツが崩れてしまうため、顔の輪郭とパーツを重点的に学習させる必要があると考えられる。そこで今後は、顔の輪郭も含めたパーツにおけるラベル情報を保持したまま学習することができるようにモデルの改良を行う。

### 参考文献

- [1] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros.  
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, 2017 IEEE International Conference on Computer Vision (ICCV), 2017.
- [2] 中島悠輔. 敵対的生成ネットワーク (GAN) を用いたイラストレータの個性を反映した似顔絵生成手法, 神奈川工科大学院情報工学専攻修士論文(未公開), 2021.
- [3] Jie Hu, Li Shen, Gang Sun. Squeeze-and-Excitation Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- [4] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, Hanqing Lu. Dual Attention Network for Scene Segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3146-3154, 2019.